

УДК 519.2

08.00.13 Математические и инструментальные методы экономики (экономические науки)

ОСНОВНЫЕ ТРЕБОВАНИЯ К СТАТИСТИЧЕСКИМ МЕТОДАМ АНАЛИЗА ДАННЫХ

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н.
профессор, РИНЦ SPIN-код: 4342-4994
prof-orlov@mail.ru

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5

Статья посвящена обоснованию полезности выработки, обсуждения и распространения системы основных требований к разработке и применению статистических методов анализа данных, к их описанию в публикациях, диссертациях и т.п. Автор в течение полувека консультировал научных работников различных специальностей, рецензировал их статьи и книги, оппонировал диссертации. Эта деятельность дала возможность познакомиться с сотнями конкретных исследований, посвященных разработке и применению статистических методов. Были выявлены разнообразные недостатки при проведении исследований и публикации их результатов, которые мешают их восприятию, а в ряде случаев ставят под сомнение адекватность выводов. Поэтому автор пришел к выводу о целесообразности выработки естественных требований к методам обработки данных и представлению результатов статистического анализа. Настоящая статья посвящена первоначальному рассмотрению ряда формулировок таких требований. Исходим из современной парадигмы прикладной статистики (основанной на непараметрической и нечисловой статистике), сменившей примитивную парадигму XIX в. и устаревшую парадигму середины XX в., основанной на использовании параметрических систем распределений. При описании и обсуждении процедур статистического анализа начинать надо с вероятностно-статистических моделей порождения изучаемых данных. Анализ многообразия моделей регрессионного анализа приводит к выводу, что не существует единой "стандартной модели". Согласно теории измерений первый шаг при анализе данных - выявление шкал, в которых они измерены. Статистические выводы должны быть инвариантны относительно допустимых преобразований шкал измерения. Поскольку распределения реальных данных ненормальны, предпочтения следует отдавать непараметрическим методам. Возможность применения параметрических семейств распределений должны быть тщательно

UDC 519.2

08.00.13 Mathematical and instrumental methods of Economics (economic sciences)

BASIC REQUIREMENTS FOR STATISTICAL METHODS OF DATA ANALYSIS

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor, RSCI SPIN-code: 4342-4994
prof-orlov@mail.ru

Bauman Moscow State Technical University, Moscow, Russia

The article is devoted to the substantiation of the usefulness of developing, discussing and disseminating a system of basic requirements for the development and application of statistical methods for data analysis, for their description in publications, dissertations, etc. For half a century, the author advised scientists of various specialties, reviewed their articles and books, and opposed dissertations. This activity provided an opportunity to get acquainted with hundreds of case studies on the development and application of statistical methods. Various shortcomings have been identified in the conduct of studies and the publication of their results, which hinder their perception, and in some cases cast doubt on the adequacy of the conclusions. Therefore, the author came to the conclusion about the expediency of developing natural requirements for data processing methods and presenting the results of statistical analysis. This article is devoted to the initial consideration of a number of formulations of such requirements. We proceed from the modern paradigm of applied statistics (based on non-parametric and non-numerical statistics), which replaced the primitive paradigm of the 19th century. and the outdated paradigm of the middle of the 20th century, based on the use of parametric distribution systems. When describing and discussing the procedures of statistical analysis, it is necessary to start with probabilistic-statistical models for generating the data under study. An analysis of the diversity of regression analysis models leads to the conclusion that there is no single "standard model". According to measurement theory, the first step in analyzing data is to identify the scales on which they are measured. Statistical inferences must be invariant under allowable transformations of measurement scales. Since the distributions of real data are non-normal, preference should be given to non-parametric methods. The possibility of using parametric families of distributions must be carefully justified. When testing statistical hypotheses, both the null and alternative hypotheses must be specified. It is necessary to study the stability of the conclusions drawn from the model with respect to acceptable changes in the initial data and assumptions of the model. Neural network

обоснована. При проверке статистических гипотез должны быть указаны как нулевая, так и альтернативная гипотезы. Необходимо изучение устойчивости выводов, получаемых на основе модели, относительно допустимых изменений исходных данных и предпосылок модели. Нейросетевые методы анализа данных являются частью прикладной статистики

methods of data analysis are part of applied statistics

Ключевые слова: МАТЕМАТИЧЕСКИЕ И СТАТИСТИЧЕСКИЕ МЕТОДЫ ЭКОНОМИКИ, АНАЛИЗ ДАННЫХ, ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКАЯ МОДЕЛЬ, ПРИКЛАДНАЯ СТАТИСТИКА, НЕПАРАМЕТРИКА, НЕЧИСЛОВАЯ СТАТИСТИКА, ТЕОРИЯ ИЗМЕРЕНИЙ, РЕГРЕССИОННЫЙ АНАЛИЗ, НЕЙРОСЕТЕВЫЕ МЕТОДЫ АНАЛИЗА ДАННЫХ

Keywords: MATHEMATICAL AND STATISTICAL METHODS OF ECONOMICS, DATA ANALYSIS, PROBABILISTIC AND STATISTICAL MODEL, APPLIED STATISTICS, NONPARAMETERS, NONNUMERICAL STATISTICS, MEASUREMENT THEORY, REGRESSION ANALYSIS, NEURAL NETWORK METHODS OF DATA ANALYSIS

<http://dx.doi.org/10.21515/1990-4665-181-026>

Introduction

It seems useful to develop, discuss and disseminate the basic requirements for the development and application of statistical methods for data analysis, for their description in publications, dissertations, etc. Why was this work necessary? It would seem that there are many textbooks, and they should be followed. However, it is often not possible to extract from textbooks and other methodological literature specific recommendations for conducting and preparing for publication of one's own work. In addition, we have to admit that publications often contain errors that have been wandering from one publication to another for decades. One of these errors is analyzed in articles [1, 2].

Constant consulting for half a century of scientific workers of various specialties, reviewing their articles and books, opposing dissertations made it possible to get acquainted with hundreds of specific studies on the development and application of statistical methods. A critical analysis of the accumulated material made it possible to develop a general approach to such studies and a number of particular methods [3, 4]. In addition, as a result of such an analysis, various shortcomings were identified in conducting research and publishing their results, which interfere with adequate perception, and in some cases cast

<http://ej.kubagro.ru/2022/07/pdf/26.pdf>

doubt on the validity of the conclusions. This substantiates our opinion that it is expedient to formulate and discuss natural requirements for data processing methods and presentation of the results of statistical analysis of specific data.

The first attempt to implement this idea was made in the recommendations [5] and the report [6]. In the same vein, Appendix 3 “Methodology for Comparative Analysis of Related Econometric Models” was prepared in [7, 8]. An attempt was made to single out the main characteristics of the methods of applied statistics and to formulate requirements for these methods (ie, for the values of the mentioned characteristics of the methods). For example, one of the requirements is that statistical inferences must be invariant under valid transformations of measurement scales.

In order to "standardize mathematical tools" (we use the terminology of N. Bourbaki [9, p. 253]), it seems appropriate to start work on the certification of statistical methods and related software packages, as well as training courses and materials [10], the rules for preparing for the publication of theoretical and practical research.

However, standardization is useful only when it is carried out by qualified specialists, otherwise, instead of benefit, we have harm. An example is the sad fate of the variety of standards for statistical methods of quality management, most of which had to be canceled due to developer errors. This situation with standardization is analyzed in detail in the article [11], and then in the textbooks [7, 8]. Obviously, the draft normative document should be subject to careful discussion based on analysis by highly qualified specialists. However, such specialists prefer to do their own research.

This article is devoted to the initial consideration of a number of formulations of requirements for data processing methods and the presentation of the results of statistical analysis of specific data. With regard to classification problems, such requirements were discussed in articles [12, 13], and the

connection with controlling - in the report [14]. We proceed from the modern paradigm of applied statistics, about which it is necessary to say a few words.

About the new paradigm of applied statistics

Statistical methods of data analysis are widely used by researchers in various fields of science. The center of this toolkit is applied statistics, i.e. the science of how to process data [3, 4]. Applications of applied statistics methods in a particular field of activity give rise to the corresponding sciences. For example, applications in economics and management (econometrics), biology (biometrics), technical research (technometrics), chemistry (chemometrics), medicine (evidence-based medicine), science and science management (scientometrics), etc.

Let's discuss the paradigm shift in applied statistics. Under the paradigm we understand the model of adequate activity in a particular field of science adopted by the most qualified core of researchers. Let us discuss the change over time in the foundations of the action model generally accepted by specialists in the field of applied statistics and methods of data analysis, more broadly - in the field of mathematical research methods.

Let's consider three paradigms actually used at the present time - primitive, obsolete, modern. Primitive corresponds to the views of the XIX and early XX centuries, obsolete - the middle of the XX century, modern - XXI century.

Let us explain by the example of the actions of modern researchers who adhere to one or another paradigm.

Starting from a primitive paradigm, naive (poorly familiar with modern applied statistics) authors use the well-known calculation formulas of the classical Student's t-test to test the statistical hypothesis that the expectation is equal to 0 without any justification and believe that they are acting correctly.

According to the outdated paradigm, at the beginning of the study, it is accepted (usually without any justification, all the more rigorous) that the measurement results have a normal distribution, then the classical Student's t-test is applied (assuming the normal distribution of the measurement results (observations, tests, analyzes, experiments) is justified).

According to the modern paradigm, nonparametric methods (based on the central limit theorem [4, 13]) should be used to test the hypothesis under consideration, since it is well known that the distributions of real data, as a rule, are not normal.

There is no doubt that the validity of statistical conclusions increases with the transition from a primitive paradigm to an outdated one and then to a modern one. Despite the progress in the development of applied statistics, all three paradigms are currently used in the practice of scientific work in various fields. Let us discuss how this affects the quality of research results and the quality of scientific publications.

We state that the primitive paradigm is the cookbook paradigm. Persons who adhere to this paradigm follow recipes compiled by someone without thinking. The use of common software products without understanding the applied methods can provoke such calculations. However, quite often the final conclusions turn out to be useful from the standpoint of the applied area. But sometimes they can be grossly erroneous. The danger of thoughtless use of software products was warned by prof. V.V. Nalimov [16], an outstanding researcher in the field of statistical methods.

From parametric statistics to non-parametric statistical methods

The outdated paradigm is the paradigm of the middle of the 20th century. The views of the beginning of the 20th century froze in it, when the first results of a new branch of science - mathematical statistics - were obtained. According to the outdated paradigm, the sample elements are considered as independent

random variables, the distributions of which are included in one or another parametric family of distributions - normal, logistic, exponential, Weibull - Gnedenko, Cauchy, Laplace, gamma distributions, beta distributions, etc. All these families are included into a four-parameter family of distributions, introduced by the founder of mathematical statistics K. Pearson at the beginning of the 20th century. In order to streamline the results of measurements (observations, analyzes, tests, experiments, surveys), he and his followers adopted a working hypothesis, that the distributions of real data always coincide with some element of his four-parameter family. Then the development of the theory of parametric mathematical statistics began, in which the problems of estimation and testing of hypotheses were solved for samples from certain parametric families. A number of remarkable mathematical models and results have been obtained, for example, related to the maximum likelihood method, Student's, Pearson's (chi-square), Fisher's, Rao-Kramer's inequality, etc. The multivariate normal distribution has proved to be very useful for the development of regression and discriminant analyses. Apparently, because the density of such a distribution at the point in which the problems of estimation and testing of hypotheses were solved for samples from certain parametric families. A number of remarkable mathematical models and results have been obtained, for example, related to the maximum likelihood method, Student's, Pearson's (chi-square), Fisher's, Rao-Kramer's inequality, etc. The multivariate normal distribution has proved to be very useful for the development of regression and discriminant analyses. Apparently, because the density of such a distribution at the point in which the problems of estimation and testing of hypotheses were solved for samples from certain parametric families. A number of remarkable mathematical models and results have been obtained, for example, related to the maximum likelihood method, Student's, Pearson's (chi-square), Fisher's, Rao-Kramer's inequality, etc. The multivariate normal distribution has proved to be very useful for the development of regression and

discriminant analyses. Apparently, because the density of such a distribution at the point Y is a quadratic form in the Y coordinates, and the regression and discriminant analysis algorithms correspond to the transformations of this quadratic form with a linear change of coordinates.

Parametric mathematical statistics is devoted to the main content of the textbooks on mathematical statistics that are common and at the present time. Unlike the primitive paradigm, there is a rigorous mathematical theory that allows, on the basis of the hypothesis that the distributions of the sample elements are included in one or another parametric family, to obtain computational algorithms and, based on them, useful practical recommendations. However, this mathematical-statistical theory has a fundamental drawback - the distributions of real data, as a rule, are not normal and do not belong to Pearson's four-parameter family at all [13]. This assertion is rigorously substantiated (see, for example, [17, 18]) and included in the textbooks [3, 4].

In applied work, sometimes they try to check the normality or, for example, the exponentiality of real data. It is often impossible to reject the normality hypothesis. But this cannot be considered as a final confirmation of the hypothesis that the distribution of the data under consideration is normal, since for the same data it is usually not possible to reject one or another hypothesis that the distribution of the data corresponds to another popular distribution. The reason for this outwardly paradoxical phenomenon is obvious - insufficient (small) sample size. For example, it is known that in order to find out which distribution the analyzed data corresponds to - normal or logistic, at

least 2500 observations are required [3, 4]. Real sample sizes are usually much smaller.

Researchers with a mathematical mindset have continued to develop parametric mathematical statistics in recent decades. In particular, relatively recently it has been found that instead of maximum likelihood estimates it is advisable to use one-step estimates, methods of confidence estimation for the parameters of the gamma distribution, etc. have been developed [3, 4]. We state that on the basis of parametric mathematical statistics, attempts have been made to solve many applied problems in specific areas of research. But in a number of cases, erroneous conclusions were obtained, although the proportion of such cases is noticeably smaller than the reliance on a primitive paradigm.

Parametric statistics has been replaced by non-parametric statistics based on a fundamentally different data generation model. In non-parametric statistics, in contrast to parametric statistics, sample elements with numerical values are assumed to have an arbitrary distribution function (in many cases, a continuity condition is also added).

The development of parametric statistics has now reached such a level that nonparametric methods can solve the same wide range of data analysis problems as parametric methods. The advantage of nonparametric statistics over parametric statistics is that there is no need to make unreasonable assumptions about the form of the distribution function.

Nonparametric statistics also have disadvantages. One of them is generated by the fact that real statistics quite often contain coincidences. The fact is that if the distribution function of the elements of the sample is continuous, as is customary in non-parametric statistics, then the probability of coincidence of two or more elements of the sample is 0. One of the reasons for the discussed contradiction is that the properties of the pragmatic numbers used to record the results of measurements (observations, tests, experiments, analyzes, surveys) differ from the properties of mathematical numbers (for

example, pragmatic numbers are written using a finite number of digits, and almost all real numbers require - in theory - an infinite number of digits). Approaches to analysis of coincidences when using non-parametric statistical methods, allowing to partially remove the contradiction under consideration [19].

In favor positive side of parametric statistics It should be noted that in some cases parametric methods make it possible to detect and preliminary study important effects of nonparametric statistics. So, it was already noted above that the distributions of real data, as a rule, are not normal. However, the mathematical apparatus in the case of normality is often simpler. According to an outdated paradigm, multivariate normal distributions are widely used in mathematical statistics. It is for such distributions that explicit formulas for various characteristics have been found in multivariate statistical analysis, primarily in regression formulations. This is due to the fact that the theory of quadratic forms in Euclidean space is deeply developed (quadratic forms are in the power of the exponent describing the density of the multidimensional normal distribution).

Modern paradigm of applied statistics and systemic fuzzy interval mathematics

The modern paradigm of applied statistics and, more broadly, of mathematical research methods is presented in [15, 21 - 27]. It is based on the application of non-parametric and non-numerical statistics methods. Brief information about nonparametric statistics is given in the previous section.

The core of applied statistics in 21st century became the statistics of non-numerical data (statistics of objects of non-numerical nature, non-numerical statistics), which allows a uniform approach to the analysis of statistical data of an arbitrary nature.

We call the modern paradigm of mathematical research methods here new, although its foundations were formed back in the 1980s, when, during

preparations for the creation of the All-Union Statistical Association (the founding congress was held in 1990 [28]), it was necessary to analyze the state and prospects of applied statistics.

We state that to date, theoretical research in applied statistics is carried out mainly in accordance with the modern paradigm. This is evidenced, for example, by the results of the analysis of articles published in the section "Mathematical Research Methods" of the journal "Industrial Laboratory. Diagnostics of Materials" in 2006 - 2015.[29]. (It should be noted here that this section is the key one in the field of theoretical work on applied statistics. Since its inception in 1962, more than a thousand papers on applied statistics have been published in it.) statistics published in this section.

Our work, devoted to the identification of a new paradigm of applied statistics, served as the basis for the creation of a new promising direction in theoretical and computational mathematics - systemic fuzzy interval mathematics. revealing one of the sides of the new paradigm. Its main idea is the transition from classical real numbers as the basis of mathematics to pragmatic numbers with a finite number of gradations, to fuzzy and interval numbers. The key publication is the monograph [31] of 2014, which aroused considerable interest among the scientific community. Its continuation is the monograph [32] devoted to the works of the authors for 2014–2021. The value of systemic fuzzy interval mathematics for mathematics in the XXI century is disclosed in articles [33, 34]. Let us point out several publications on this new promising direction of theoretical and computational mathematics, which we consider as the basis of mathematics of the 21st century [35 - 37].

We have to state that at present a significant proportion of applied work is carried out in the tradition of outdated or even primitive paradigms. It is inappropriate to indiscriminately deny such works. They may be useful in specific areas. However, it is indisputable that the transition to the modern

paradigm of applied statistics will raise the scientific level of research, and will also provide important results in specific areas. Unfortunately, many researchers involved in data analysis, including developers of software products on this subject, are not familiar enough with non-parametric and non-numerical statistics [30]. It is necessary to disseminate information about the modern paradigm of applied statistics more widely.

Reliance on the approaches and results of non-parametric and non-numerical statistics is one of the main requirements for statistical methods of data analysis. Let's expand on this statement.

The role of probabilistic-statistical data models

The first stage in the development and application of applied statistics methods is the selection and justification of probabilistic-statistical data models.

When describing, applying and discussing certain procedures for analyzing statistical data, attention is usually focused on the calculation formulas. Indeed - without formulas it is impossible to carry out calculations. However, the calculation algorithms are based on probabilistic-statistical models for generating the data under study. It is with these models that one must begin - both when conducting a study and when describing it.

For example, in works on applied statistics, naive authors usually understand a sample as a finite sequence of numbers. Qualified researchers in most cases use the most common sampling model, according to which the measurement results are considered as a finite sequence of realizations of independent identically distributed random variables [3, 4], modeling the results of measurements (observations, tests, experiments, analyzes, surveys).

If the general distribution function of these random variables is arbitrary, then it is necessary to turn to the methods of nonparametric statistics. For real data, coincidences of results are quite common. Therefore, in such cases, deviations from the nonparametric model are observed. As noted above, a model

for analyzing coincidences in calculating the values of nonparametric rank statistics was proposed in [19]. The interval data statistics as a component of non-numerical statistics was created to handle rounded and matched data [3, 4].

Note the persistence of prejudices. As already noted, the notions corresponding to the outdated paradigm of applied statistics that the distribution function of measurement results belongs to one of the popular families of continuous distribution functions - normal, logarithmically normal, exponential, Weibull-Gnedenko, gamma distributions, beta distributions, and etc. For samples from such families, methods for estimating parameters and testing statistical hypotheses were developed and studied in the last millennium. This set of methods has firmly taken its place in textbooks on probability theory and mathematical statistics, executed in the spirit of an outdated paradigm.

However, parametric statistics is also developing, but outdated views are stable. For example, the use of the maximum likelihood method is still promoted, although one-step estimators have just as good properties as maximum likelihood estimates. In some cases, the system of maximum likelihood equations does not have an explicit solution in the form of final calculation formulas, and it is recommended to find the corresponding estimates using one or another iterative method. Their convergence is generally not studied, although there are examples in which the lack of convergence has been demonstrated. Meanwhile, one-step estimates are calculated by finite formulas, without any iterations [3, 4].

The inclination of theorists in the field of mathematical and applied statistics to use is noticeable. multivariate normal distributions. It is for such distributions that explicit formulas for various characteristics have been found in multivariate statistical analysis, primarily in regression. According to our expert assessment, the reason is that such theorists manage to use the well-developed theory of quadratic forms in linear algebra.

It has long been established that the distributions of almost all real data are not normal (Gaussian). This statement is well substantiated experimentally with a thorough analysis of the results of measurements of various quantities [17, 18]. They also put forward theoretical arguments to justify the use of the normal distribution. Thus, it is argued that the dependence of the value of a random variable on many factors entails normality. Sometimes they increase the validity of such a judgment by adding that the factors are independent and comparable in magnitude to random variables. However, closeness to the normal distribution can only be expected if the additive data generation model is correct, when the factors are added (this statement follows from the Central Limit Theorem of probability theory). If the random variable is formed by multiplication (multiplicative data generation model), then its distribution is (asymptotically) logarithmically normal, not normal. If the model of the "weakest" link (or "strongest", record) is valid, i.e. the value of a random variable is equal to the extreme member of the variational series of factor values (respectively, the minimum or maximum), then we have the Weibull-Gnedenko distribution in the limit. This fact was established by B.V. Gnedenko in the 1940s, which explains the name of the family of distributions under consideration (although the family itself was previously used by W. Weibull). The development of statistical methods is traced in the monograph [38]. If the model of the "weakest" link (or "strongest", record) is valid, i.e. the value of a random variable is equal to the extreme member of the variational series of factor values (respectively, the minimum or maximum), then we have the Weibull-Gnedenko distribution in the limit. This fact was established by B.V. Gnedenko in the 1940s, which explains the name of the family of distributions under consideration (although the family itself was previously used by W. Weibull). The development of statistical methods is traced in the monograph [38]. If the model of the "weakest" link (or "strongest", record) is valid, i.e. the value of a random variable is equal to the extreme member of the variational

series of factor values (respectively, the minimum or maximum), then we have the Weibull-Gnedenko distribution in the limit. This fact was established by B.V. Gnedenko in the 1940s, which explains the name of the family of distributions under consideration (although the family itself was previously used by W. Weibull). The development of statistical methods is traced in the monograph [38]. which explains the name of the family of distributions under consideration (although the family itself was previously used by W. Weibull). The development of statistical methods is traced in the monograph [38]. which explains the name of the family of distributions under consideration (although the family itself was previously used by W. Weibull). The development of statistical methods is traced in the monograph [38]. which explains the name of the family of distributions under consideration (although the family itself was previously used by W. Weibull). The development of statistical methods is traced in the monograph [38].

Using a model based on the family of normal distributions can be compared to searching under a bright lantern for keys lost in dark bushes. Obviously, it's easier to look under the lantern. You can show activity. However, it is impractical to hope for a favorable outcome of the search for keys.

It follows from the above methodological analysis that it is necessary to use non-parametric models for the distributions of measurement results. Note that the possible values for the measurement results, as a rule, have a priori minimum and maximum (for example, corresponding to the limits of the scale, fixed in the technical passport of the measuring instruments). In other words, the distributions are finite. Consequently, all moments of the considered random variables exist, and their sample analogs can be used in calculations. This remark allows us to get rid of some necessary conditions in the limiting theory of mathematical statistics.

From the foregoing, the following requirement for statistical methods of data processing follows: if, for any reason, the researcher wishes to apply a parametric family of distributions, its use must be carefully justified by testing the statistical hypothesis of agreement with both the considered family and alternative families.

The role of probabilistic-statistical models in multivariate statistical analysis

Let's start with one of the main sections of multivariate statistical analysis - with regression analysis. Several basic types of regression models are used. Let's discuss the simplest setting - one independent variable and one independent variable. Let us briefly characterize the main models used.

Often, least squares models with a deterministic independent variable and parametric dependence (linear, quadratic, etc.) are used. It is natural to assume that the distribution of deviations is arbitrary (i.e., consider a non-parametric model). The derivation of the limit distributions of the parameter estimates and the regression dependence is based on the Central Limit Theorem and the linearization theorem[3, 4].

A fundamentally different type of models is based on a selection of random vectors. In most cases, the dependence is parametric, the parameters are estimated from sample data. It is natural to assume that the distribution of a two-dimensional vector is arbitrary. It is possible to talk about the estimation of the variance of an independent variable (unlike the dependent one) only in a model based on a sample of random vectors, as well as about the coefficient of determination as a criterion for the quality of the model, otherwise fundamental errors are possible [39].

Another type of regression analysis models, based on a sample of random vectors- non-parametric regression, in which both the dependence and the deviations from it are non-parametric. Dependence (as a conditional mean) is estimated using nonparametric estimates of the distribution density of a random vector.

Another option is a model in which the trend is linear, and the periodic and random components and deviations from them are nonparametric. It is intermediate between the two just discussed.

In models of the following type, there are small errors both in the values of the dependent variable and in the values of the independent variable. It is natural to describe the values of variables by intervals. In the last century, this section of applied statistics, dedicated to models of this type, was called confluent analysis; now it is included in the statistics of interval data [3, 4, 31, 40, 41].

Further development of the above classification of regression analysis models is possible. So, it is usually accepted that errors (errors, residuals) are independent identically distributed random variables. It is possible to waive both the requirement of the same distribution and the requirement of independence.

So, if the root-mean-square error is proportional to the measured value, then we come to the need to minimize not the sum of squared differences between the values of the dependent variable and the function of the independent variable, but another optimization criterion. Namely, the quotients of dividing the indicated quantities by the values of the function of the values of the independent variable are squared. In other words, in the method of least squares, it is necessary to replace absolute deviations with relative ones[42].

Instead of the sum of squared deviations, you can use other formulations of the optimization problem, for example, to minimize the sum of deviation modules (the method of least modules) or the maximum (modulo) deviation (the minimax method).

Problems adjoining regression analysis smoothing time series and statistics of random processes, in which deviations from the function of time are dependent (in contrast to regression analysis, in which such deviations are independent random variables). In other words, when modeling time series, it is quite natural to abandon the requirement of independence of errors. Moreover, since the dependence between the values of the random function on time, as a rule, decreases with increasing distance between the measurement moments,

then the independence of the errors can be postulated only when the measurement moments differ significantly from each other.

It is possible to describe errors not by random variables, but by fuzzy numbers, a special case of which are intervals, which have already been discussed above.

We do not attempt to describe all the various formulations of regression analysis. This requires monographs like [43]. However, the above brief analysis of the variety of regression analysis models leads to the conclusion that there is no single "standard model" [44]. Therefore, when solving and describing the problem of recovering dependence, it is necessary to start with the choice and justification of one or another probabilistic-statistical model for generating data.

Measurement theory as a basis for building probabilistic-statistical models

According to modern views, when conducting a statistical analysis of data, it is necessary to proceed from the theory of measurements [3, 4, 31, 40, 41]. According to this theory, the first step in analyzing data is to identify the scales on which they are measured. The main requirement is that the statistical methods used must correspond to the scales in which the data are measured.

Let's take an example. Statistical inferences based on the calculation of averages must be invariant under the allowable transformations of the measurement scales of statistical data. It is proved that for data measured on an ordinal scale, only a finite number of functions from the measurement results, namely, members of the variation series, can be used as average values. With an odd sample size - the median, and with an even - the left median or the right median. The use of, for example, the arithmetic mean or geometric mean is unacceptable. As a consequence, since ranks or scores commonly used in applied research are usually measured on an ordinal scale, the arithmetic mean cannot be calculated for them. In particular, [44].

The main requirement is that statistical inferences based on the calculation of certain statistics (functions of measurement results) must be invariant with respect to acceptable transformations of data measurement scales. Consequently, researchers in the field of the theory of applied statistics face the primary task: for each scale they use, find out which data analysis algorithms from the family of algorithms they consider can be used in this scale. The conclusions regarding the use of the Cauchy family of means are briefly described above.

The inverse problem is also important - for a certain data analysis algorithm to find out in which scale it can be used. It has been found that Pearson's linear pair correlation coefficient corresponds to the interval scale, while Spearman's and Kendall's nonparametric rank correlation coefficients are aimed at studying the relationship of ordinal variables.

Based on the theory of measurements, let us briefly consider a fairly well-known method of analyzing hierarchies. The input data in this method are the results of paired comparisons, they are measured in ordinal scales. And the results of calculations by the method of analysis of hierarchies are expressed in the scale of intervals, according to the enthusiasts of this method. From the point of view of measurement theory, this is unacceptable. The results of the calculations (statistical inferences) must be measured on the same scale as the original data. Therefore, from the point of view of measurement theory, the method of analysis of hierarchies should not be used. We recommend using adequate methods for analyzing expert assessments, in particular, the methods of arithmetic mean ranks, median ranks, and matching of clustered rankings [46, 47].

Training samples in problems of diagnostics and neural networks

When discussing the ideas and results of this article, it was noted that it is quite natural to extend the developed requirements to an adjacent (closely related) area - neural network data processing. Given the significant

interpenetration of probabilistic-statistical and neural network methods, this seems to be very useful.

In our opinion, we should start with a discussion of terminology [48]. How do applied statistics and neural network methods compare?

In order to implement this idea, let us consider, as a basic example, the relationship between applied statistics and neural network data processing in the field of mathematical classification theory [49]. There are three sections in this theory - the construction of classifications, the study of classifications, the application of classifications [3, 4]. If the study of classifications is usually considered part of the statistics of non-numerical data, then the other two areas have very different names in the literature.

Synonyms for the concept of "building classifications", in our opinion, are as follows: cluster analysis (the full form of the term is cluster analysis), pattern recognition without a teacher, typology, taxonomy, grouping, classification without a teacher, dichotomy ... The author came to this conclusion in the result of the analysis of hundreds of papers using the listed terms.

Similarly, synonyms for the term "application of classifications" are: methods of discrimination (discriminant analysis), in one of the most common variant - mathematical methods of diagnostics, pattern recognition with a teacher, automatic classification with a teacher, statistical classification ...

Here, the "teacher" is understood as the methods of constructing decision rules based on training samples. It is assumed that for each of the classes there is a training sample, i.e. a selection of elements from this class. Based on the training samples, a decision rule is built on which class to attribute the newly arriving object to.

When we talk about unsupervised algorithms, this means that we are talking about building a classification based on the analysis of data from a single training sample, for the elements of which it is not indicated which class this

element belongs to. Unsupervised algorithms are based on certain measures of proximity between elements (indicators of difference).

Currently, "neural networks" is a very popular term. We are talking about various mathematical models (as well as algorithms developed on their basis, their software or hardware implementation), built by analogy with the networks of nerve cells of a living organism. The first such models were developed in the middle of the 20th century. in the study of processes occurring in the human brain. An attempt was made to model these processes (at the level of knowledge of that time). It is now known that the human brain works differently, suggest neural network enthusiasts.

With a careful analysis of the main ideas of neural network methods, it becomes obvious that these models are primarily intended for solving classification problems based on the analysis of training samples. In other words, the classical problems of classification theory are not solved in the same way as they were done earlier in applied statistics.

The theory of mathematical statistics makes it possible to compare classification algorithms by quality. For diagnostic problems, it is advisable to carry out a comparison based on the predictive power of the algorithm [50, 51]. It turns out that neural network algorithms, as a rule, are not optimal. For example, it has been proven in classification theory that for assigning a newly arriving object to one of two classes given by training samples, the decision rule based on nonparametric estimates of the probability distribution densities corresponding to the classes is (asymptotically) optimal [3, 4, 49]. Neural network methods cannot give a better result than this decision rule. Unfortunately,

We come to the conclusion that neural networks, pattern recognition methods, and, for example, genetic algorithms are other names for a number of long-developed sections of applied statistics (statistical methods for data analysis) [52, 53]. Through the efforts of journalists and publicists who are not

very versed in the ideas and scientific results of applied statistics, the new terminology has become the center of attention of the scientific community. This happened for non-scientific reasons, which are revealed in the final part of the article [48].

conclusions

This article substantiates the need to develop a system of requirements for statistical models and methods in their creation, application and teaching, including their description in scientific and methodological publications.

We emphasize that, first of all, a probabilistic-statistical model of data generation should be presented and justified. A useful analysis of the hierarchical structure of the concept of "model" and potential sources of errors in the construction, study, application and teaching of probabilistic-statistical models of real data is presented in the article [54].

Let us briefly formulate a number of requirements for statistical methods analyzed above.

Since almost all distributions of real data are non-normal, preference should be given to non-parametric formulations. The possibility of using parametric families of distributions must be carefully justified.

In accordance with the theory of testing statistical hypotheses, not only the null hypothesis, but also the alternative one should be indicated, only then it is possible to discuss the power of the criterion.

It is necessary to study the stability of the conclusions obtained on the basis of the organizational and economic model with respect to acceptable changes in the initial data and model assumptions [55]. In particular, statistical inferences must be invariant under allowable scale transformations.

Substantiation of the main requirements for statistical methods of data analysis on the example of classification problems (diagnostics and cluster analysis) is devoted to the corresponding sections of works [3, 4, 52, 53].

A number of further publications will be devoted to the problems of developing a system of requirements for statistical models and methods. The author is grateful to Prof. IN. Tolcheev for helpful remarks.

Литература

1. Орлов А.И. Распространенная ошибка при использовании критериев Колмогорова и омега-квадрат // Заводская лаборатория. 1985. Т.51. №1. С. 60-62.
2. Орлов А.И. Непараметрические критерии согласия Колмогорова, Смирнова, Омега-квадрат и ошибки при их применении // Научный журнал КубГАУ. 2014. №97. С. 647–675.
3. Орлов А. И. Прикладная статистика. - М.: Экзамен, 2006. - 671 с.
4. Орлов А. И. Прикладной статистический анализ. — М.: Ай Пи Ар Медиа, 2022. — 812 с. <https://www.iprbookshop.ru/117038.html>, <https://doi.org/10.23682/117038>
5. Орлов А. И. Рекомендации. Прикладная статистика. Методы обработки данных. Основные требования и характеристики / А.И. Орлов, Н.Г. Миронова, В.Н. Фомин, А.Н. Черчинцев. - М.: ВНИИСтандартизации, 1987. - 62 с.
6. Орлов А. И. Основные характеристики статистических методов обработки данных и требования к ним / А.И. Орлов, Н.Г. Миронова, В.Н. Фомин, А.Н. Черчинцев // Доклады Московского Общества испытателей природы 1987 г. Общая биология: Морфология и генетика процессов роста и развития. - М.: Наука, 1989. С.66-68.
7. Орлов А. И. Эконометрика. Учебник для вузов. Изд. 3-е, переработанное и дополненное. - М.: Изд-во «Экзамен», 2004. – 576 с.
8. Орлов А. И. Эконометрика : учебное пособие. — М., Саратов : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 676 с.
9. Бурбаки Н. Очерки по истории математики. - М.: Изд-во иностранной литературы, 1963. – 292 с.
10. Орлов А. И. Сертификация статистических методов, пакетов программ и систем обучения // Международная конференция по интервальным и стохастическим методам в науке и технике (ИНТЕРВАЛ-92). Москва, 22-26 сентября 1992 г. Сборник трудов. – М.: Изд-во МЭИ, 1992. - Том 1. С. 125-128. Т.2. С. 88-88.
11. Орлов А. И. Сертификация и статистические методы (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 1997. Т.63. №3. С. 55-62.
12. Орлов А. И. Основные требования к методам анализа данных (на примере задач классификации) // Научный журнал КубГАУ. 2020. №159. С. 239–267. <http://dx.doi.org/10.21515/1990-4665-159-017>
13. Орлов А. И. Основные требования к математическим методам классификации // Заводская лаборатория. Диагностика материалов. 2020. Т.86. № 11. С. 67-78.
14. Орлов А. И. Контроллинг и статистические методы / Контроллинг в экономике, организации производства и управлении: сборник научных трудов X международного конгресса по контроллингу, (Ярославль, 22 октября 2021 г.) / Под научной редакцией д.э.н., профессора С.Г. Фалько / НП «Объединение контроллеров». – М.: НП «Объединение контроллеров», 2021. – С. 65 - 74.

15. Орлов А. И. Новая парадигма математических методов исследования // Заводская лаборатория. Диагностика материалов. 2015. Т.81. №7. С. 5-5.
16. Налимов В. В. Теория эксперимента. - М.: Наука, 1971. - 208 с.
17. Орлов А.И. Часто ли распределение результатов наблюдений является нормальным? // Заводская лаборатория. 1991 Т.57. №7. С. 64-66.
18. Орлов А.И. Распределения реальных статистических данных не являются нормальными // Научный журнал КубГАУ. 2016. №117. С. 71 – 90.
19. Орлов А. И. Модель анализа совпадений при расчете непараметрических ранговых статистик // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №11. С. 66-72.
20. Орлов А. И. Оценивание размерности вероятностно-статистической модели / Научный журнал КубГАУ. 2020. №162. С. 1–36.
21. Орлов А.И. Новая парадигма прикладной статистики // Статистика и прикладные исследования: сборник трудов Всерос. научн. конф. – Краснодар: Издательство КубГАУ, 2011. – С. 206-217.
22. Орлов А. И. Новая парадигма прикладной статистики // Заводская лаборатория. Диагностика материалов. 2012. Т.78. №1. С. 87-93.
23. Орлов А.И. Новая парадигма математической статистики // Материалы республиканской научно-практической конференции «Статистика и её применения – 2012». Под редакцией проф. А.А. Абдушукурова. – Ташкент: НУУз, 2012. – С. 21-36.
24. Орлов А.И. Основные черты новой парадигмы математической статистики // Научный журнал КубГАУ. 2013. №90. С. 188-214.
25. Орлов А.И. О новой парадигме прикладной математической статистики // Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. / Перм. гос. нац. иссл. ун-т. – Пермь, 2013. – Вып. 25. –С. 162-176.
26. Орлов А.И. О новой парадигме математических методов исследования // Научный журнал КубГАУ. 2016. №122. С. 807–832.
27. Орлов А. И. Смена парадигм в прикладной статистике // Заводская лаборатория. Диагностика материалов. 2021. Т.87. № 7. С. 6-7.
28. Орлов А.И. Создана единая статистическая ассоциация // Вестник Академии наук СССР. 1991. №7. С. 152-153.
29. Орлов А. И. Развитие математических методов исследования (2006 – 2015 гг.) // Заводская лаборатория. Диагностика материалов. 2017. Т.83. №1. Ч.1. С. 78-86.
30. Орлов А.И. Статистические пакеты – инструменты исследователя // Заводская лаборатория. Диагностика материалов. 2008. Т.74. №5. С.76-78.
31. Орлов, А. И. Системная нечеткая интервальная математика - основа математики XXI века / А. И. Орлов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – 2021. – № 165. – С. 111-130. – DOI 10.21515/1990-4665-165-011. – EDN BAPGPX.
32. Орлов, А. И. Анализ данных, информации и знаний в системной нечеткой интервальной математике / А. И. Орлов, Е. В. Луценко. – Краснодар : Кубанский государственный аграрный университет им. И.Т. Трубилина, 2022. – 405 с. – ISBN 978-5-907550-62-9. – DOI 10.13140/RG.2.2.15688.44802. – EDN OQULUW.
33. Орлов А.И. Системная нечеткая интервальная математика - основа математики XXI века // Научный журнал КубГАУ. 2021. №165. С. 111–130.
34. Орлов А.И. Системная нечеткая интервальная математика - основа инструментария математических методов исследования // Заводская лаборатория. Диагностика материалов. 2022. Т.88. №7. С. 5-7. DOI: <https://doi.org/10.26896/1028-6861-2022-88-7-5-7>. <https://www.elibrary.ru/item.asp?id=49182008>

35. Орлов А.И., Луценко Е.В. О развитии системной нечеткой интервальной математики // *Философия математики: актуальные проблемы. Математика и реальность. Тезисы Третьей всероссийской научной конференции; 27-28 сентября 2013 г.* / Редкол.: Бажанов В.А. и др. – Москва, Центр стратегической конъюнктуры, 2013. – С. 190–193.

36. Орлов, А. И. Системная нечеткая интервальная математика (сним) – перспективное направление теоретической и вычислительной математики / А. И. Орлов, Е. В. Луценко // *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета.* – 2013. – № 91. – С. 163-215. – EDN RKNKEZ.

37. Луценко, Е. В. Когнитивные функции как обобщение классического понятия функциональной зависимости на основе теории информации в АСК-анализе и системной нечеткой интервальной математике / Е. В. Луценко, А. И. Орлов // *Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета.* – 2014. – № 95. – С. 58-81. – EDN RVEYDF.

38. Лойко, В. И. Высокие статистические технологии и системно-когнитивное моделирование в экологии / В. И. Лойко, Е. В. Луценко, А. И. Орлов. – Краснодар : Кубанский государственный аграрный университет имени И.Т. Трубилина, 2019. – 258 с. – ISBN 978-5-00097-855-9. – EDN PJGBXC.

39. Орлов А. И. Ошибки при использовании коэффициентов корреляции и детерминации // *Заводская лаборатория. Диагностика материалов.* 2018. Т.84. № 3. С. 68-72.

40. Орлов А.И. Теория принятия решений. Учебник для вузов. — М.: Экзамен, 2006. — 576 с.

41. Орлов А.И. Теория принятия решений : учебник. — М.: Ай Пи Ар Медиа, 2022. — 826 с. — ISBN 978-5-4497-1467-1. — Текст : электронный // IPR SMART : [сайт]. — URL: <https://www.iprbookshop.ru/117047.html>

42. Копаев Б.В. В методе наименьших квадратов надо заменить абсолютные отклонения относительными // *Заводская лаборатория. Диагностика материалов.* 2012. Т.88. №7. С. 76-76.

43. Себер Дж. Линейный регрессионный анализ. - М.: Мир, 1980. - 456 с.

44. Орлов А. И. Многообразие моделей регрессионного анализа (обобщающая статья) / *Заводская лаборатория. Диагностика материалов.* 2018. Т.84. №5. С. 63-73.

45. Орлов А.И. Характеризация средних величин шкалами измерения // *Научный журнал КубГАУ.* 2017. №134. С. 877–907.

46. Орлов А. И. Организационно-экономическое моделирование: в 3 ч. Ч.2. Экспертные оценки. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2011. - 486 с.

47. Орлов А. И. Искусственный интеллект: экспертные оценки. — М.: Ай Пи Ар Медиа, 2022. — 436 с. <https://doi.org/10.23682/117030>

48. Орлов А. И. Смена терминологии в развитии науки // *Научный журнал КубГАУ.* 2022. №177. С. 232–246. <http://dx.doi.org/10.21515/1990-4665-177-013>

49. Орлов А. И. Базовые результаты математической теории классификации // *Научный журнал КубГАУ.* 2015. №110. С. 219–239.

50. Орлов А.И. Прогностическая сила как показатель качества алгоритма диагностики // *Статистические методы оценивания и проверки гипотез: межвуз. сб. науч. тр. Вып.23.* – Пермь: Перм. гос. нац. иссл. ун-т, 2011. – С. 104-116.

51. Орлов А.И. Прогностическая сила – наилучший показатель качества алгоритма диагностики // *Научный журнал КубГАУ.* 2014. №99. С. 15–32.

52. Орлов А. И. Искусственный интеллект: нечисловая статистика. — М.: Ай Пи Ар Медиа, 2022. — 446 с. <https://www.iprbookshop.ru/117028.html>,

<https://doi.org/10.23682/117028>

53. Орлов А. И. Искусственный интеллект: статистические методы анализа данных. — М.: Ай Пи Ар Медиа, 2022. — 843 с. <https://www.iprbookshop.ru/117029.html>, <https://doi.org/10.23682/117029>

54. Савельев О. Ю. Модель: иерархия понятия и потенциальный источник ошибок // Инновации в менеджменте. 2021. №28. С. 54-58.

55. Орлов А. И. Устойчивые экономико-математические методы и модели : монография. — М.: Ай Пи Ар Медиа, 2022. — 337 с. <https://www.iprbookshop.ru/117049.html>, <https://doi.org/10.23682/117049>

References

1. Orlov A.I. Rasprostrannaya oshibka pri ispol`zovanii kriteriev Kolmogorova i omega-kvadrat // Zavodskaya laboratoriya. 1985. T.51. №1. S. 60-62.

2. Orlov A.I. Neparаметрические критерии согласия Колмогорова, Смирнова, Омга-квadrat и oshibki pri ix primenenii // Nauchny`j zhurnal KubGAU. 2014. №97. S. 647–675.

3. Orlov A. I. Prikladnaya statistika. - M.: E`kzamen, 2006. - 671 s.

4. Orlov A. I. Prikladnoj statisticheskij analiz. — M.: Aj Pi Ar Media, 2022. — 812 с. <https://www.iprbookshop.ru/117038.html>, <https://doi.org/10.23682/117038>

5. Orlov A. I. Rekomendacii. Prikladnaya statistika. Metody` obrabotki danny`x. Osnovny`e trebovaniya i xarakteristiki / A.I. Orlov, N.G. Mironova, V.N. Fomin, A.N. Charchincev. - M.: VNIStandartizacii, 1987. - 62 s.

6. Orlov A. I. Osnovny`e xarakteristiki statisticheskix metodov obrabotki danny`x i trebovaniya k nim / A.I. Orlov, N.G. Mironova, V.N. Fomin, A.N. Charchincev // Doklady` Moskovskogo Obshhestva ispy`tatelej prirody` 1987 g. Obshhaya biologiya: Morfologiya i genetika processov rosta i razvitiya. - M.: Nauka, 1989. S.66-68.

7. Orlov A. I. E`konometrika. Uchebnik dlya vuzov. Izd. 3-e, pererabotannoe i dopolnennoe. - M.: Izd-vo «E`kzamen», 2004. – 576 s.

8. Orlov A. I. E`konometrika : uchebnoe posobie. — M., Saratov : Internet-Universitet Informacionny`x Texnologij (INTUIT), Aj Pi Ar Media, 2020. — 676 с.

9. Burbaki N. Oчерки по истории математики. - M.: Izd-vo inostranoj literatury`, 1963. – 292 s.

10. Orlov A. I. Sertifikaciya statisticheskix metodov, paketov programm i sistem obucheniya // Mezhdunarodnaya konferenciya po interval`ny`m i stoxasticheskim metodam v nauke i texnike (INTERVAL-92). Moskva, 22-26 sentyabrya 1992 g. Sbornik trudov. – M.: Izd-vo ME`I, 1992. - Tom 1. S. 125-128. T.2. S. 88-88.

11. Orlov A. I. Sertifikaciya i statisticheskie metody` (obobshhayushhaya stat`ya) // Zavodskaya laboratoriya. Diagnostika materialov. 1997. T.63. №3. S. 55-62.

12. Orlov A. I. Osnovny`e trebovaniya k metodam analiza danny`x (na primere zadach klassifikacii) // Nauchny`j zhurnal KubGAU. 2020. №159. S. 239–267. <http://dx.doi.org/10.21515/1990-4665-159-017>

13. Orlov A. I. Osnovny`e trebovaniya k matematicheskim metodam klassifikacii // Zavodskaya laboratoriya. Diagnostika materialov. 2020. T.86. № 11. S. 67-78.

14. Orlov A. I. Kontrolling i statisticheskie metody` / Kontrolling v e`konomie, organizacii proizvodstva i upravlenii: sbornik nauchny`x trudov X mezhdunarodnogo kongressa po kontrollingu, (Yaroslavl`, 22 oktyabrya 2021 g.) / Pod nauchnoj redakciej d.e`n., professora S.G. Fal`ko / NP «Ob`edinenie kontrollerov». – M.: NP «Ob`edinenie kontrollerov», 2021. – S. 65 - 74.

15. Orlov A. I. Novaya paradigma matematicheskix metodov issledovaniya // Zavodskaya laboratoriya. Diagnostika materialov. 2015. T.81. №.7. S. 5-5.

<http://ej.kubagro.ru/2022/07/pdf/26.pdf>

16. Nalimov V. V. Teoriya e`ksperimenta. - M.: Nauka, 1971. - 208 s.
17. Orlov A.I. Chasto li raspredelenie rezul'tatov nablyudenij yavlyayetsya normal'ny`m? // Zavodskaya laboratoriya. 1991 T.57. №7. S. 64-66.
18. Orlov A.I. Raspredeleniya real'ny`x statisticheskix dannyx ne yavlyayutsya normal'ny`mi // Nauchny`j zhurnal KubGAU. 2016. №117. S. 71 – 90.
19. Orlov A. I. Model` analiza sovpadenij pri raschete neparametricheskix rangovy`x statistik // Zavodskaya laboratoriya. Diagnostika materialov. 2017. T.83. №11. S. 66-72.
20. Orlov A. I. Ocenivanie razmernosti veroyatnostno-statisticheskoy modeli / Nauchny`j zhurnal KubGAU. 2020. №162. S. 1–36.
21. Orlov A.I. Novaya paradigma prikladnoj statistiki // Statistika i prikladny`e issledovaniya: sbornik trudov Vseros. nauchn. konf. – Krasnodar: Izdatel'stvo KubGAU, 2011. – S. 206-217.
22. Orlov A. I. Novaya paradigma prikladnoj statistiki // Zavodskaya laboratoriya. Diagnostika materialov. 2012. T.78. №1. S. 87-93.
23. Orlov A.I. Novaya paradigma matematicheskoy statistiki // Materialy` respublikanskoj nauchno-prakticheskoy konferencii «Statistika i eyo primeneniya – 2012». Pod redakciej prof. A.A. Abdushukurova. – Tashkent: NUUZ, 2012. – S. 21-36.
24. Orlov A.I. Osnovny`e cherty` novoj paradigmy` matematicheskoy statistiki // Nauchny`j zhurnal KubGAU. 2013. №90. S. 188-214.
25. Orlov A.I. O novoj paradigme prikladnoj matematicheskoy statistiki // Statisticheskie metody` ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. / Perm. gos. nacz. issl. un-t. – Perm`, 2013. – Vy`p. 25. –S. 162-176.
26. Orlov A.I. O novoj paradigme matematicheskix metodov issledovaniya // Nauchny`j zhurnal KubGAU. 2016. №122. S. 807–832.
27. Orlov A. I. Smena paradigmi v prikladnoj statistike // Zavodskaya laboratoriya. Diagnostika materialov. 2021. T.87. № 7. S. 6-7.
28. Orlov A.I. Sozdana edinaya statisticheskaya asociaciya // Vestnik Akademii nauk SSSR. 1991. №7. S. 152-153.
29. Orlov A. I. Razvitie matematicheskix metodov issledovaniya (2006 – 2015 gg.) // Zavodskaya laboratoriya. Diagnostika materialov. 2017. T.83. №1. Ch.1. S. 78-86.
30. Orlov A.I. Statisticheskie pakety` – instrumenty` issledovatelya // Zavodskaya laboratoriya. Diagnostika materialov. 2008. T.74. №5. S.76-78.
31. Orlov, A. I. Sistemnaya nechetkaya interval`naya matematika - osnova matematiki XXI veka / A. I. Orlov // Politematicheskij setevoy e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2021. – № 165. – S. 111-130. – DOI 10.21515/1990-4665-165-011. – EDN BAPGPX.
32. Orlov, A. I. Analiz dannyx, informacii i znaniy v sistemnoj nechetkoj interval`noj matematike / A. I. Orlov, E. V. Lucenko. – Krasnodar : Kubanskij gosudarstvenny`j agrarny`j universitet im. I.T. Trubilina, 2022. – 405 s. – ISBN 978-5-907550-62-9. – DOI 10.13140/RG.2.2.15688.44802. – EDN OQULUW.
33. Orlov A.I. Sistemnaya nechetkaya interval`naya matematika - osnova matematiki XXI veka // Nauchny`j zhurnal KubGAU. 2021. №165. S. 111–130.
34. Orlov A.I. Sistemnaya nechetkaya interval`naya matematika - osnova instrumentariya matematicheskix metodov issledovaniya // Zavodskaya laboratoriya. Diagnostika materialov. 2022. T.88. №7. S. 5-7. DOI: <https://doi.org/10.26896/1028-6861-2022-88-7-5-7>. <https://www.elibrary.ru/item.asp?id=49182008>
35. Orlov A.I., Lucenko E.V. O razvitii sistemnoj nechetkoj interval`noj matematiki // Filosofiya matematiki: aktual'ny`e problemy`. Matematika i real'nost`. Tezisy` Tret`ej vsrossijskoj nauchnoj konferencii; 27-28 sentyabrya 2013 g. / Redkol.: Bazhanov V.A. i dr. – Moskva, Centr strategicheskoy kon`yunktury`, 2013. – S. 190–193.

36. Orlov, A. I. Sistemnaya nechetkaya interval'naya matematika (snim) – perspektivnoe napravlenie teoreticheskoy i vy`chislitel`noj matematiki / A. I. Orlov, E. V. Lucenko // Politematicheskij setевой e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2013. – № 91. – S. 163-215. – EDN RKNKEZ.

37. Lucenko, E. V. Kognitivny`e funkcii kak obobshhenie klassicheskogo ponyatiya funkcional`noj zavisimosti na osnove teorii informacii v ASK-analize i sistemnoj nechetkoj interval`noj matematike / E. V. Lucenko, A. I. Orlov // Politematicheskij setевой e`lektronny`j nauchny`j zhurnal Kubanskogo gosudarstvennogo agrarnogo universiteta. – 2014. – № 95. – S. 58-81. – EDN RVEYDF.

38. Lojko, V. I. Vy`sokie statisticheskie texnologii i sistemno-kognitivnoe modelirovanie v e`kologii / V. I. Lojko, E. V. Lucenko, A. I. Orlov. – Krasnodar : Kubanskij gosudarstvenny`j agrarny`j universitet imeni I.T. Trubilina, 2019. – 258 s. – ISBN 978-5-00097-855-9. – EDN PJGBXC.

39. Orlov A. I. Oshibki pri ispol`zovanii koefficientov korrelyacii i determinacii // Zavodskaya laboratoriya. Diagnostika materialov. 2018. T.84. № 3. S. 68-72.

40. Orlov A.I. Teoriya prinyatiya reshenij. Uchebnik dlya vuzov. — M.: E`kzamen, 2006. — 576 s.

41. Orlov A.I. Teoriya prinyatiya reshenij : uchebnik. — M.: Aj Pi Ar Media, 2022. — 826 c. — ISBN 978-5-4497-1467-1. — Tekst : e`lektronny`j // IPR SMART : [sajt]. — URL: <https://www.iprbookshop.ru/117047.html>

42. Kopaev B.V. V metode naimen`shix kvadratov nado zamenit` absolyutny`e otkloneniya otnositel`ny`mi // Zavodskaya laboratoriya. Diagnostika materialov. 2012. T.88. №7. S. 76-76.

43. Seber Dzh. Linejny`j regressionny`j analiz. - M.: Mir, 1980. - 456 s.

44. Orlov A. I. Mnogoobrazie modelej regressionnogo analiza (obobshhayushhaya stat`ya) / Zavodskaya laboratoriya. Diagnostika materialov. 2018. T.84. №5. S. 63-73.

45. Orlov A.I. Xarakterizaciya srednix velichin shkalami izmereniya // Nauchny`j zhurnal KubGAU. 2017. №134. S. 877–907.

46. Orlov A. I. Organizacionno-e`konomicheskoe modelirovanie: v 3 ch. Ch.2. E`kspertny`e ocenki. - M.: Izd-vo MGTU im. N. E`. Baumana, 2011. - 486 s.

47. Orlov A. I. Iskusstvenny`j intellekt: e`kspertny`e ocenki. — M.: Aj Pi Ar Media, 2022. — 436 c. <https://doi.org/10.23682/117030>

48. Orlov A. I. Smena terminologii v razvitii nauki // Nauchny`j zhurnal KubGAU. 2022. №177. S. 232–246. <http://dx.doi.org/10.21515/1990-4665-177-013>

49. Orlov A. I. Bazovy`e rezul`taty` matematicheskoy teorii klassifikacii // Nauchny`j zhurnal KubGAU. 2015. №110. S. 219–239.

50. Orlov A.I. Prognosticheskaya sila kak pokazatel` kachestva algoritma diagnostiki // Statisticheskie metody` ocenivaniya i proverki gipotez: mezhvuz. sb. nauch. tr. Vy`p.23. – Perm`: Perm. gos. nacz. issl. un-t, 2011. – S. 104-116.

51. Orlov A.I. Prognosticheskaya sila – nailuchshij pokazatel` kachestva algoritma diagnostiki // Nauchny`j zhurnal KubGAU. 2014. №99. S. 15–32.

52. Orlov A. I. Iskusstvenny`j intellekt: nechislovaya statistika. — M.: Aj Pi Ar Media, 2022. — 446 c. <https://www.iprbookshop.ru/117028.html>, <https://doi.org/10.23682/117028>

53. Orlov A. I. Iskusstvenny`j intellekt: statisticheskie metody` analiza danny`x. — M.: Aj Pi Ar Media, 2022. — 843 c. <https://www.iprbookshop.ru/117029.html>, <https://doi.org/10.23682/117029>

54. Savel`ev O. Yu. Model`: ierarxiya ponyatiya i potencial`ny`j istochnik oshibok // Innovacii v menedzhmente. 2021. №28. S. 54-58.

55. Orlov A. I. Ustojchivye e`konomiko-matematicheskie metody` i modeli : monografiya. — M.: Aj Pi Ar Media, 2022. — 337 c.
<https://www.iprbookshop.ru/117049.html>, <https://doi.org/10.23682/117049>